

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 239 406 A3

(12)

EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
16.03.2005 Bulletin 2005/11

(51) Int Cl.7: **G06K 9/68, G06K 9/72**

(43) Date of publication A2:
11.09.2002 Bulletin 2002/37

(21) Application number: **02004377.4**

(22) Date of filing: **05.03.2002**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
 MC NL PT SE TR**
 Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **07.03.2001 JP 2001063968**

(71) Applicants:
 • **Kabushiki Kaisha Toshiba**
Tokyo (JP)
 • **Suzuki, Masakazu**
Fukuoka-shi, Fukuoka-ken (JP)

(72) Inventors:
 • **Suzuki, Masakazu, Kyushu University**
Fukuoka-shi, Fukuoka-ken (JP)
 • **Yokota, Kazuaki**
2016, Shimmachi, 9-chome, Ome-shi, Tokyo (JP)
 • **Eto, Yuko, A910, Kureare Toshiba Ome**
Ome-shi, Tokyo (JP)

(74) Representative: **HOFFMANN - EITLE**
Patent- und Rechtsanwälte
Arabellastrasse 4
81925 München (DE)

(54) **Device and method for character recognition and for recognition of mathematical expressions**

(57) A mathematical expression recognizing device comprises a character recognition unit (112) configured to recognize characters in a document image containing a text and a mathematical expression, a first dictionary (201) configured to store a pair of evaluation scores for each type of word that can be identified by means of normal expression, the score showing the possibility of belonging to the text and that of belonging to the mathematical expression, an evaluation unit (113) configured to obtain the evaluation scores showing the possibility of belonging to the text and that of belonging to the math-

ematical expression for each of the words included in the characters recognized by the character recognition unit with reference to the first dictionary, and a mathematical expression detecting unit (114) configured to search for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, thereby detecting characters belonging to the mathematical expression.

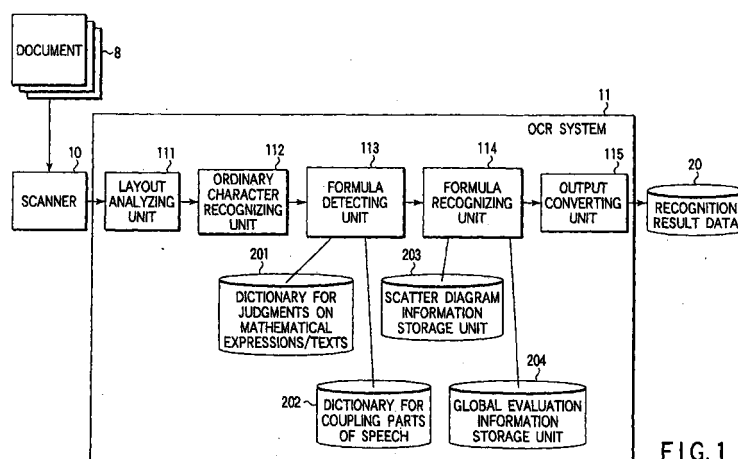


FIG. 1

EP 1 239 406 A3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 00 4377

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	SRIHARI R K ET AL: "Incorporating syntactic constraints in recognizing handwritten sentences" IJCAI-93. PROCEEDINGS OF THE THIRTEENTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE MORGAN KAUFMANN PUBLISHERS SAN MATEO, CA, USA, vol. 2, 28 August 1993 (1993-08-28), pages 1262-1267, XP002314334 CHAMBERY, FRANCE	1,2,9, 12,15	G06K9/68 G06K9/72
Y	* abstract * * page 1262, left-hand column, paragraph 3 - page 1263, right-hand column, paragraph 2 * * page 1263, left-hand column, paragraph 2 *	3,16	
Y	----- BLOSTEIN D ET AL: "COMPUTING WITH GRAPHS AND GRAPH TRANSFORMATIONS" SOFTWARE PRACTICE & EXPERIENCE, JOHN WILEY & SONS LTD. CHICHESTER, GB, vol. 29, no. 3, March 1999 (1999-03), pages 197-217, XP000802322 ISSN: 0038-0644 * abstract; figures 7-10 * * page 204, paragraph 3 - page 210, paragraph 1 * ----- -/--	3,7,8, 10,11, 13,16	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06K
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 24 January 2005	Examiner Müller, M
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons _____ & : member of the same patent family, corresponding document	

3
EPO FORM 1503 03 02 (P04C01)



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 00 4377

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	TWAAKYONDO H M ET AL: "Structure analysis and recognition of mathematical expressions" DOCUMENT ANALYSIS AND RECOGNITION, 1995., PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON MONTREAL, QUE., CANADA 14-16 AUG. 1995, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, vol. 1, 14 August 1995 (1995-08-14), pages 430-437, XP010230970 ISBN: 0-8186-7128-9 * abstract * * page 431, left-hand column *	3,7,8, 10,11, 13,16	
A	CHANG CHAO-HUANG: "Word class discovery for postprocessing chinese handwriting recognition" PROC. 15TH INT. CONF. ON COMPUTATIONAL LINGUISTICS - COLING 1994, 5 August 1994 (1994-08-05), pages 1221-1225, XP002314335 * abstract * * page 1222, right-hand column, paragraph 3 - paragraph 5 * * page 1223, right-hand column, paragraph 5 - paragraph 6 * ----- -/--	1,2,9, 12,15	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 24 January 2005	Examiner Müller, M
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ----- & : member of the same patent family, corresponding document</p>			

3
EPO FORM 1503 03 82 (P04C01)



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 00 4377

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	TOUMIT J -Y ET AL: "A hierarchical and recursive model of mathematical expressions for automatic reading of mathematical documents" PROC. OF THE 5TH INT. CONF. ON DOCUMENT ANALYSIS AND RECOGNITION. ICDAR '99 (CAT. NO.PR00318) IEEE COMPUT. SOC LOS ALAMITOS, CA, USA, 20 September 1999 (1999-09-20), pages 119-122, XP002314336 BANGALORE, INDIA ISBN: 0-7695-0318-7 * abstract * * page 119, right-hand column, paragraph 1 - paragraph 4 *	1,2,9, 12,15	
A	JÖRG HUNSINGER ET AL.: "A single-stage top-down probabilistic approach towards understanding spoken and handwritten mathematical formulas" PROC. 6TH INT. CONF. ON SPOKEN LANGUAGE PROCESSING - CHINA MILITARY FRIENDSHIP PUBLISHERS, vol. 4, 16 October 2000 (2000-10-16), pages 386-389, XP002314337 BEIJING, CHINA * abstract; figures 2,3 * * page 386, section 1 * * page 387, right-hand column *	3,7,8, 10,11, 13,16	
The present search report has been drawn up for all claims			TECHNICAL FIELDS SEARCHED (Int.Cl.7)
Place of search Munich		Date of completion of the search 24 January 2005	Examiner Müller, M
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>			

3
EPO FORM 1503 03.82 (P04C01)